1950-08-14

A comparison of admitted and non-admitted groups at the U.S. Naval Postgraduate School

Mollenkopf, William G.

Princeton, New Jersey; Education Testing Service

RB-50-53

RESEARCH BULLETIN

A COMPARISON OF ADMITTED AND NON-ADMITTED GROUPS AT THE U. S. NAVAL POSTGRADUATE SCHOOL

A Report to the Superintendent of
the U. S. Naval Postgraduate School

William G. Mollenkopf

8-14-50

Educational Testing Service

A Comparison of Admitted and Non-Admitted

Groups at the U. S. Naval Postgraduate School

A B S T R A C T

In the fall of 1948 the five-part Naval Postgraduate School Aptitude Test Form WNPF was administered at shore installations and on ships to applicants for admission to the Postgraduate School. Test data for those later admitted to the School have now been compared with similar data for those who were not selected, to allow comparisons of the applicant and admitted groups and to estimate the effects of the present selection process on the magnitudes of the test validities being obtained in the experimental test-development program being carried on at the Postgraduate School.

It was found that the greatest reduction in range of talent had occurred for the Mathematics Aptitude part of the test, and, as a consequence, that the validity of this part was most markedly affected by the selection process. Lesser effects were noted for the Reading Comprehension, Spatial Intersections, and Physics parts of the test. No reduction in range of talent was observed for the Verbal Antonyms part, and since it seems probable that the validity of this test is not seriously affected by the selection process, its observed low validity may be concluded not to be a result of rigorous selection on verbal ability.

It was concluded that in the future interpretation of test validities, the fact must be taken into account that the present method of selection restricts the range of ability in mathematics aptitude somewhat more than in other areas. Observed validities for mathematics tests are therefore likely to have been lowered more by the selection process than those of other tests involved in the study.

A Comparison of Admitted and Non-Admitted Groups at the U. S. Naval Postgraduate School

Introduction

In May 1948 the Educational Testing Service began a research program for the Navy which had as its objective the development of a battery of tests that could be used as an effective part of the process of selecting students for admission to the U. S. Naval Postgraduate School. A battery of 20 tests was assembled and administered in July 1948 to the incoming student officers at the School.

However, it was realized at the time that adequate information regarding characteristics of the applicant group could not be gained by testing only those individuals who were admitted to the School, since the individuals who enter the School each year constitute a selected group taken from a considerably larger number of candidates. To the extent that the existing methods of selection identified individuals who did well in the School, and rejected those who would have done poorly, the test results for the selected group would differ from the results for the total group of applicants. As one important consequence, validity coefficients of tests might be markedly lowered by any great reduction in the range of talent.

Consequently, the proposal was made that a specially prepared experimental test be given to all the officers who applied for admission to the Postgraduate School in July 1949, for the purpose of comparing applicant and admitted groups and especially to permit estimates of how the present selection process affected the range of talent in each of several areas. This proposal was approved by the authorities at the Postgraduate School, who in turn secured the approval and cooperation of the Bureau of Naval Personnel.

Procedure

By force of circumstances the testing of applicants for admission had to be carried out at whatever shore installation or on whatever ship these individuals were stationed. This had several implications regarding what kind of test and what testing conditions would be suitable:

a. The testing time required should be short, say 3 hours at the most.

b. The test should be a power test. Speeding of material would not be practicable.

c. The test should be practically self-administering. The procedure of testing should involve little supervision, except for an over-all time limit and security measures to safeguard test content.

d. Some particular day or days should probably be designated, on which all applicants would take the test.

e. All test books and answer sheets should be sent out and returned to a central point.

When approval of the proposal was given, the Bureau of Naval Personnel agreed to act as the forwarding and receiving agency for the test materials. Arrangements were made that the test materials would be sent to the commanding officer of the applicant, who would then see that the test was given at the proper time and under suitable conditions, and afterward return the test materials to the Bureau.

Content of the Test. In considering what types of material might best be included in the test book, several criteria for judgment seemed to be important. These were

a. Likelihood of being a good predictor of academic success at the Postgraduate School.

b. Dissimilarity with other tests to be included (low intercorrelations of measures being thought desirable).

c. Acceptability to the group tested. (Apparent validity was felt to be important in determining motivation.)

d. Suitability for use under unspeeded conditions.

e. Similarity with tests in the battery already administered in July 1948.

On the basis of these considerations the decision was made to construct a test having five parts, as follows:

Part 1, Reading Comprehension. This test consisted of six passages of scientific reading material each of which was followed by a set of five questions to be answered on the basis of what was stated or implied in the passage. There were thus 30 questions in all in this part of the test.

Part 2, Mathematics Aptitude. This was a mathematical aptitude test consisting of 30 items chosen so as to minimize the effects on the score of length and recency of formal training in mathematics.

Part 3, Verbal Antonyms. This part contained 30 verbal antonyms items similar to those used in the verbal section of the College Board Scholastic Aptitude Test.

Part 4, Spatial Intersections. Each item in this test showed a drawing of a solid figure cut by a plane. The student was called upon to select from five figures presented him the one which correctly represented the shape on the cutting plane of the intersection between the solid figure and this plane. There were 50 items in the part.

Part 5, Elementary Physics. The items in this part were on topics covered in the usual high-school course in physics. There were 30 items in the part.

Administration of the Test. As had been planned, the test books and answer sheets were sent by the Bureau of Naval Personnel directly to the applicant's commanding

officer, together with a covering letter and a form containing both instructions for administration of the test and a certificate to be completed by the commanding officer in which he stated when and to whom he had administered the test, and that he had carried out the instructions.

Most of the tests were administered during the week of November 15-19, 1948. While many of the test books were returned within the next few weeks, a few -- chiefly coming from far distant places -- were not returned until the end of January, 1949. Eleven more individuals were given the test at the Postgraduate School itself in July, 1949. In all, the Postgraduate School Aptitude Test was administered to 394 persons.

Analysis of the Data. During the scoring of the answer sheets, one paper was found to carry the notation that the test book had been defective. Of the other 393 answer sheets, 242 were for officers not admitted to the Postgraduate School and 151 were for officers admitted to the School.

Since the scores were to be used in considering the effects of the selection process on the validities of the tests, the admitted group was broken down into those pursuing a three-year curriculum and those in the one-year Applied Communications curriculum. (Validation studies have been based on the former group.) There were 135 officers in the first of these groups and 16 in the second. For four of the 135 officers, either no grades or grades for only one term were later received from the Postgraduate School. Five other persons did not attempt any of the items in the last section of the test. Hence the basic Admitted group in the analysis contained 126 individuals.

Of the Non-Admitted group of 242, six persons did not attempt any of the items in the last section of the test. The basic Non-Admitted group consequently contained 236 individuals.

In Table 1 are presented the means and standard deviations for these Admitted and Non-Admitted groups. For all parts except Part 3, Verbal Antonyms, the difference between the means of the Admitted and Non-Admitted groups is sufficiently large that a difference so great would be expected to occur by chance less than one time in every hundred such instances.

Since some drop-out did occur on the test because not all individuals had time to try every item, the results for Part 5 may be open to some question. The results in Table 1 are for those persons who tried some items in the last part. As a matter of further interest, the data for persons trying at least the first 80 per cent of the items (i.e., 24 of the 30) in Part 5 were examined. From Table 2 it can be seen that for these persons the difference between the means of the Admitted and Non-Admitted groups is slightly larger than for the groups compared in Table 1. Elimination from consideration of those individuals who tried less than 80 per cent of the items in Part V decreased the standard deviation of the scores for both groups, as might have been expected. When a significance test was applied to the data in Table 2, it was again found that the likelihood that the observed difference in means would have occurred by chance was less than 1 in 100.

In order to determine what effect the selective process might have had on the validity coefficients obtained from study of the selected group, the variabilities of the Admitted and Applicant (i.e., Admitted plus Non-Admitted) groups must be compared. In Table 3 are presented the standard deviations of Admitted and Applicant groups. To assess the amount of restriction which has occurred in the range of talent on the various parts of the test as a consequence of the selection, the ratio of the two standard deviations was calculated for each part.

The most notable result is that the Applicant and Admitted groups are not greatly different in variability on most of the parts. The highest degree of re-

striction has occurred on Part 2, the Mathematics Aptitude section of the test. In view of the kinds of material included and the functions tested by the various parts of the test, the fact that the greatest amount of change in range of talent occurred on the mathematics aptitude variable is not at all surprising.

Table 1

Means and Standard Deviations for Admitted and Non-Admitted Groups on Each of the Five Parts of the Naval Postgraduate School Aptitude Test WNPF

Part	Admitted Group* N = 126		Non-Admitted Group N = 236	
1. Reading Comprehension	M σ	21.14 4.61	M σ	19.67 5.13
2. Mathematics Aptitude	M σ	19.64 4.52	M σ	16.62 5.29
3. Verbal Antonyms	M σ	18.35 5.89	M σ	18.11 5.63
4. Spatial Intersections	M σ	37.23 9.53	M σ	34.09 10.55
5. Elementary Physics	M σ	18.43 6.83	M σ	16.30 6.78

*For the Admitted group, only cases completing at least the first two terms of a three-year curriculum are included.

Table 2

Means and Standard Deviations on the Physics Part for Persons Trying at Least the First 80 Per Cent of the Items

	Admitted Group N = 107	Non-Admitted Group N = 198
Mean	20.27	17.97
Standard Deviation	5.32	5.79

Table 3

Standard Deviations of Admitted and Applicant Groups and Ratios of These

Part	Std. Dev. Admitted Group	Std. Dev. Applicant Group	Ratio Adm./App.
1. Reading Comprehension	4.61	5.00	0.92
2. Mathematics Aptitude	4.52	5.24	0.86
3. Verbal Antonyms	5.89	5.73	1.03
4. Spatial Intersections	9.53	10.31	0.92
5. Physics	6.83	6.87	0.99
5a. Physics (those trying 80 per cent or more of the items)	5.32	5.74	0.93

In the present situation, selection of students was made by a number of Navy boards employing as criteria of judgment various pieces of evidence on each applicant. (The test-score data were not available to the Boards.) If it could be assumed that selection was made on a single composite measure, into which a man's previous academic grades, record of performance of duty, recommendations, etc. were entered in a definite and systematic way, the data on this composite measure could be used to estimate accurately how the validity coefficients should be corrected for the effects of selection.

Since this information is not and cannot be made available, it seems legitimate to provide estimates based on reasonable assumptions. First let us assume that the restriction which has occurred in the range of scores on each part is the same as would have resulted if the test score itself had been directly used to select students. The table below shows the observed validities for the five parts of WNPF and the validities after correction using this assumption.*

The corrected validities in Table 4 are in most instances not appreciably higher than the observed coefficients, though the six-point increase for the Mathematics Aptitude Test is worth noting. It is to be mentioned that making the assumption we have used in this instance (i.e., that selection was directly on the test) is generally considered to lead to an under-estimate of the correction under these circumstances.

Dr. Harold Gulliksen, Research Adviser at the Educational Testing Service, has provided a formula for which we have much of the needed information. This when

---

*The table in Appendix E-4 of Personnel Research and Test Development in the Bureau of Naval Personnel, edited by Stuit, gives a convenient way of finding the corrected values.

Table 4

Validities as Corrected Assuming
Direct Selection on Test Part

Part	Observed Validity	Corrected Validity Assuming Direct Selection on Part
1. Reading Comprehension	.37	.40
2. Mathematics Aptitude	.46	.52
3. Verbal Antonyms	.23	--*
4. Spatial Intersections	.44	.47
5. Physics	.48	.48
5a. Physics (those trying 80 per cent or more of the items)	.50	.53

*Since the variability actually increased slightly in selection, the formula does not apply, and hence no corrected value is reported.

modified to meet the demands of this particular problem, is

$$R_{yz} = \frac{r_{yz}s_y\sqrt{R_{xz}^2(s_y^2 - S_y^2) + S_y^2R_{xy}^2} - R_{xz}(s_y^2 - S_y^2)}{R_{xy}S_y^2}$$

in which

x is the variable on which direct selection occurs,

y and z are variables on which incidental selection occurs,

R_{xy}, R_{xz} are correlations of the incidental-selection variable with the direct-selection variable for the unrestricted group,

R_{yz} is the correlation between the two incidental-selection variables for the unrestricted group,

r_{yz} is the correlation between the two incidental-selection variables for the restricted group,

S_y is the standard deviation of one incidental-selection variable for the unrestricted group, and

s_y is the standard deviation of this same incidental-selection variable for the restricted group.

In the present situation, x is the selection variable which we have hypothesized, y is the test, and z is the criterion (average grades).

The correlation between selection variable and test, R_{xy}, can be estimated if the assumption is made that the selection variable was a normally distributed measure on which candidates were rigidly chosen, all those above some critical point being accepted and all others rejected. Under these circumstances the necessary data are at hand for computing a biserial coefficient of correlation between the hypothetical selection-variable and each of the tests. These correlations when computed were found to be as follows: Part 1, .18; Part 2, .35; Part 3, .03; Part 4, .19; Part 5 (all candidates), .19; Part 5 (candidates finishing 80 per cent of items), .25. The fact that Part 2, Mathematics Aptitude, was found to have the highest correlation with the selection variable is consistent with results already presented showing that the greatest restriction of range occurred on this variable.

All of the quantities involved in the right-hand side of the equation are now available except R_{xz}, the validity of the selection variable. Various values of this validity were assumed, and the test validities then estimated by use of the formula, with the results presented in Table 5.

It is at once apparent that for assumed validities of the hypothetical selection variable of .20 and higher, the corrected values in Table 5 are greater than

those given in Table 4. Considerably higher values are then found for the validity of the Mathematics Aptitude part than before correction, and this test is indicated to be the most valid of the five. No corrections were applicable to the Verbal Antonyms validity, since the assumptions involved in the correction formula were clearly not met inasmuch as the variability of the Admitted group exceeded that of the Applicant group. The validities of the other parts all were increased appreciably if for Part 5 the data for those trying at least 80 per cent of the items are considered rather than the data for the entire group, which are of doubtful meaning.

Table 5

Test Validities as Corrected Using the Modified Formula, for Various Assumed Validity of the Selection Variable

Part of Test	Observed Validity	Corrected Validity Assuming Validity of Selection Variable for Applicant Group Is						
		.00	.10	.20	.30	.40	.50	.60
1. Reading Comprehension	.37	.34	.42	.48	.53	.51	--*	--*
2. Mathematics Aptitude	.46	.41	.47	.53	.58	.62	.64	.64
3. Verbal Antonyms	.23	--**	--**	--**	--**	--**	--**	--**
4. Spatial Intersections	.44	.41	.48	.54	.56	.56	--*	--*
5. Physics	.48	.48	.48	.48	.49	.49	.48	.48
5a. Physics (those trying 80 per cent or more of the items)	.50	.46	.52	.56	.58	.60	.59	.53

*Assumed validities of the selection variable as great as these led to a negative quantity under the radical in the equation, and hence the result is an imaginary number.

**As for Table 4, since the variability increased slightly in selection, the formula did not apply.

Conclusions

Comparison of scores on the five parts of the Naval Postgraduate School Aptitude Test made by the Admitted and Non-Admitted groups has shown that for four of the five parts the mean scores for the Admitted Group were significantly higher than those for the Non-Admitted Group. For the Verbal Antonyms part the mean scores did not differ significantly.

When the variability of the Admitted group on each of the five parts of the test was compared with the variability of the total Applicant group, it was found that the greatest amount of reduction in range of talent from Applicant to selected group had occurred for the Mathematics Aptitude part. The amounts of reduction in range of talent were not large for the other parts.

To make correction-for-selection formulas strictly applicable, it would be necessary to have complete information on the actual selection variable or variables. Since this information was not and could not be available, several assumptions were made which permitted use of correction formulas. Study of the results obtained when the observed validity coefficients were corrected under these conditions indicates that of the five variables the one whose validity has been most markedly affected by the present selection process is Mathematics Aptitude.

Previous reports have indicated that of the various types of measure tried and evaluated to date, those showing highest validity for predicting Postgraduate School grades have been tests in mathematics, physics, engineering, and reading comprehension. The present study involved tests in three of these four areas, but no test in engineering. Of these three, the evidence of this study indicates that the validities of the mathematics tests have probably been lowered most by the selection process, and that the validities for reading comprehension, spatial intersections, and physics were also lowered appreciably.

The main implication of the results is that in future interpretation of obtained validities, it will be advisable to take into account the fact that the presently used method of selecting students restricts the range of talent in mathematics somewhat more sharply than it does the ranges of ability in certain other areas. In future comparisons of the validities of two tests, allowance should be made for this differential effect. A mathematics test and verbal test having the same observed validity would not have the same corrected validity -- the mathematics test will apparently have had its validity much more affected by the selection process, and for equal observed validity must have had appreciably higher initial validity.